

CRUSH: The NSI Data Compression Utility

377

Ed Seiler (STX/GSFC Code 935)
February 13, 1991

N91-27032

CRUSH : The NSI Data Compression Utility

CRUSH is a data compression utility that provides the user with several lossless compression techniques available in a single application. CRUSH was originally developed for the NSSDC as a result of requests for such a package from a users working group meeting. It is intended that the future development of CRUSH will depend upon feedback from the user community to identify new features and capabilities desired by the users.

CRUSH provides an extension to the UNIX Compress program and the various VMS implementations of Compress that many users are familiar with. An important capability added by CRUSH is the addition of additional compression techniques and the option of automatically determining the best technique for a given data file.

The CRUSH software is written in C and is designed to run on both VMS and UNIX systems. VMS files that are compressed will regain their full file characteristics upon decompression. To the extent possible, compressed files can be transferred between VMS and UNIX systems, and thus be decompressed on a different system than they were compressed on.

Version 1 of CRUSH is currently available from the NSSDC. This version is a VAX VMS implementation. Version 2, which has the full range of capabilities for both VMS and UNIX implementations, will be available shortly. A VMS Backup file containing the source of version 1 is available at NSSDCB::ANON_DIR:[PUBLIC]CRUSH.BCK, and equivalently via anonymous FTP at NSSDCA.GSFC.NASA.GOV. It is anticipated that version 2 will be made available through the NSINIC. Watch for an announcement soon.

CRUSH has been developed as part of the research of data compression techniques for the Configurable High-Rate Processor project at NASA Goddard Space Flight Center. Edward Seiler, of ST Systems Corp. is the developer of the software, and can be contacted via E-mail at SEILER@AMARNA.GSFC.NASA.GOV

CRUSH

The NSI Data Compression Utility

Features:

- Compresses VMS and UNIX files, and keeps VMS file characteristics upon decompression
- Can compress single files or entire directories
- Extends the capabilities of the CMPR and COMPRESS programs
- (Currently) provides 3 different methods, or automatic selection of the method that provides the best compression. More methods available soon
- Compressed files can be recognized either from a file extension with `_c` appended, or by finding ASCII “CRUSHED” in first 7 bytes
- Will decompress files that were compressed by UNIX COMPRESS

Usage Considerations:

- Except for ASCII headers, compressed files are binary, so they cannot be printed
- Three methods available:
 - Lempel-Ziv-Welch (LZW) : just as used in CMPR, it's relatively fast
 - Witten-Neal-Cleary (WNC) : uses arithmetic coding
 - Adaptive WNC : usually slower than LZW but often compresses better
- Automatic "method" tries each method for a file (or a sample section of a large file) and chooses the best one, but this may take a while
- Decompressor figures out which method to use by itself

Usage

CRUSH filename(s)

- a single filename, a list of filenames, and wildcards may be used
- user will be prompted if filename omitted
- if file is named “myfile.dat”, compressed file will be named “myfile.dat_c”

/METHOD = { lzw | wnc | adap | auto }

- lzw : Lempel-Ziv-Welch
- wnc : Witten-Neal-Cleary
- adap : Adaptive WNC
- auto : all of the above are tried

/DELETE

- delete “myfile.dat” after “myfile.dat_c” is created

/OUTPUT = filespec

- CRUSH /OUTPUT=[mydir.dir] `.*` compresses all files and directs them to the subdirectory [mydir.dir]
- CRUSH/OUTPUT=TT: directs uncrushed output to the terminal

UNCRUSH filename

- for a crushed file named “myfile.dat_c”, generates the uncrushed file “myfile.dat”

382

/OUTPUT = filespec

- UNCRUSH/OUTPUT=goofy.name myfile.dat_c renames the uncrushed file to “goofy.name”